KEY_Lesson24_Basic_Stats_IV_Significance

December 10, 2021

1 Lesson 24: Statistical Significance

1.1 An introduction to hypothesis testing

Statistical significance is one of the most important parts of statistics—it's what allows us to make conclusions about our data.

But first, what does **statistical significance** mean? - **Definition: statistical significance** is the likelihood that the difference between a given variation and the baseline is not due to random chance. Statistical significance is calculated using different mathematical formulas, which we'll see later in the lesson.

We can determine if our difference is **statistically significant** by comparing our calculated significance value to the significance level. - **Definition: statistical significance level** is the level at which one can accept whether an event or difference is statistically significant, or not due to random chance. This term is denoted as α. The customary significance level is 5%. - **Definition: confidence level** is the opposite of significance level, where the confidence level indicates the degree of confidence that the result did not occur by chance. This term is calculated as $1 - \alpha$. The customary confidence level is 95%. - **For example**, if you run an A/B testing experiment with a **significance level (α) of 5%**, this means that if you determine a winner, you can be **95% confident** $(1 - \alpha)$ that the observed results are real and not an error caused by randomness. But there is a **5%** chance that you could be wrong.

Next, when is statistical significance most **practically used**? - It is used in statistical hypothesis testing. - **For example**, you want to know whether or not having a healthier diet will result in lower levels of C-Reactive protein, and hence fewer incidents of infection and inflammation.

What is **hypothesis testing**? - **Definition: hypothesis testing** is the use of statistic to determine the probability that a given hypothesis is true. - There are two types of statistical hypotheses. - **Definition**: The **null hypothesis**, denoted by H_o, is usually the hypothesis that sample observations result purely from chance. The most common null hypothesis is that the variable in question is equal to 0, i.e. this indicates that the variable has zero effect on the outcome of interest. - **Definition**: The **alternative hypothesis**, denoted by H_1 or H_a, is the hypothesis that sample observations are influenced by some non-random cause. A common alternative hypothesis is that the variable in question has a non-zero effect on the outcome.

2 What is a t-test?

Let's apply these definitions to our diet and C-Reactive Protein example. We want to know whether or not having a healthier diet will result in lower levels of C-Reactive protein, and hence fewer

incidents of infection and inflammation. - What is the **null hypothesis**? - **Answer**: Having a healthier diet will have zero effect on the levels of C-Reactive protein.

- What is the **alternative hypothesis**?
    - **Answer**: Having a healthier diet will lower the levels of C-Reactive protein.

There are so many statistical tests that are appropriate for use if certain assumptions are met. But for this lesson, we will focus on the Student's t-test. - Here a link to all statistical tests: https://machinelearningmastery.com/statistical-hypothesis-tests-in-python-cheat-sheet/.

Let's start with a **Student's t-test**. This test will assess whether samples from two independent population provide evidence that the populations have different means.

Another aspect of hypothesis testing is the "use of a statistic". **What is this statistic (test statistic)?** - **Definition**: A **test statistic** is a quantity derived from the sample and is used in statistical hypothesis testing to determine whether the null hypothesis should be accepted or rejected. - There are a wide variety of test statistics to use for a given problem. But as we are using a Student's t-test, we will focus on the **t-statistic**.

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2/N_1 + s_2^2/N_2}}$$

where N_1 and N_2 are the sample sizes, $\bar{x}_1$ and $\bar{x}_2$ are the sample means, and s_1^2 and s_2^2 are the sample variances

Now that we have calculated our t-statistic, what do we do with it to reject or fail to reject the null hypothesis? - We need to compare our t-statistic to a critical value. - **Definition**: a **critical value** is a point on the test distribution that is compared to the test statistic to determine whether to fail to reject or reject the null hypothesis.

What is the critical value in the t-distribution when conducting a Student's t-test?
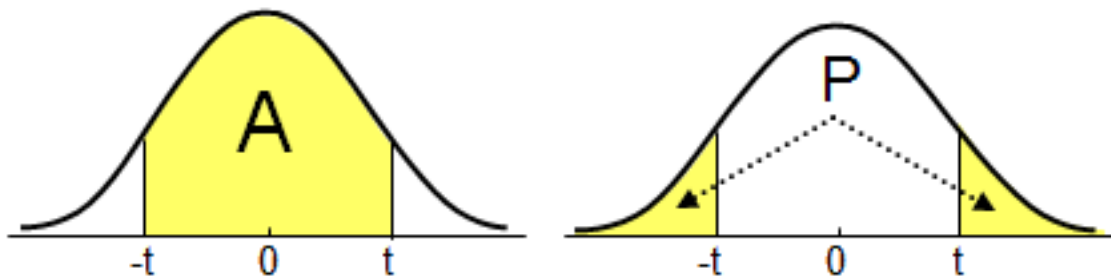
DF	A P	0.8 0.2	0.9 0.1	0.95 0.05	0.98 0.02	0.99 0.01	0.995 0.005	0.998 0.002	0.999 0.001
1		3.078	6.314	12.706	31.82	63.657	127.321	318.309	636.619
2		1.886	2.92	4.303	6.965	9.925	14.089	22.327	31.599
3		1.638	2.353	3.182	4.541	5.841	7.453	10.215	12.924
4		1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.61
5		1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6		1.44	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7		1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8		1.397	1.86	2.306	2.897	3.355	3.833	4.501	5.041
9		1.383	1.833	2.262	2.821	3.25	3.69	4.297	4.781
10		1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11		1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12		1.356	1.782	2.179	2.681	3.055	3.428	3.93	4.318
13		1.35	1.771	2.16	2.65	3.012	3.372	3.852	4.221
14		1.345	1.761	2.145	2.625	2.977	3.326	3.787	4.14
15		1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16		1.337	1.746	2.12	2.584	2.921	3.252	3.686	4.015
17		1.333	1.74	2.11	2.567	2.898	3.222	3.646	3.965
18		1.33	1.734	2.101	2.552	2.878	3.197	3.61	3.922
19		1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20		1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.85
21		1.323	1.721	2.08	2.518	2.831	3.135	3.527	3.819
22		1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23		1.319	1.714	2.069	2.5	2.807	3.104	3.485	3.768
24		1.318	1.711	2.064	2.492	2.797	3.09	3.467	3.745
25		1.316	1.708	2.06	2.485	2.787	3.078	3.45	3.725
26		1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27		1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.69
28		1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29		1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30		1.31	1.697	2.042	2.457	2.75	3.03	3.385	3.646
31		1.309	1.695	2.04	2.453	2.744	3.022	3.375	3.633
32		1.309	1.694	2.037	2.449	2.738	3.015	3.365	3.622
33		1.308	1.692	2.035	2.445	2.733	3.008	3.356	3.611
34		1.307	1.691	2.032	2.441	2.728	3.002	3.348	3.601
35		1.306	1.69	2.03	2.438	2.724	2.996	3.34	3.591
36		1.306	1.688	2.028	2.434	2.719	2.991	3.333	3.582
37		1.305	1.687	2.026	2.431	2.715	2.985	3.326	3.574
38		1.304	1.686	2.024	2.429	2.712	2.98	3.319	3.566
39		1.304	1.685	2.023	2.426	2.708	2.976	3.313	3.558
40		1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
42		1.302	1.682	2.018	2.418	2.698	2.963	3.296	3.538
44		1.301	1.68	2.015	2.414	2.692	2.956	3.286	3.526
46		1.3	1.679	2.013	2.41	2.687	2.949	3.277	3.515
48		1.299	1.677	2.011	2.407	2.682	2.943	3.269	3.505
50		1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60		1.296	1.671	2	2.39	2.66	2.915	3.232	3.46
70		1.294	1.667	1.994	2.381	2.648	2.899	3.211	3.435
80		1.292	1.664	1.99	2.374	2.639	2.887	3.195	3.416
90		1.291	1.662	1.987	2.369	2.632	2.878	3.183	3.402
100		1.29	1.66	1.984	2.364	2.626	2.871	3.174	3.391
120		1.289	1.658	1.98	2.358	2.617	2.86	3.16	3.373
150		1.287	1.655	1.976	2.351	2.609	2.849	3.145	3.357
200		1.286	1.652	1.972	2.345	2.601	2.839	3.131	3.34
300		1.284	1.65	1.968	2.339	2.592	2.828	3.118	3.323
500		1.283	1.648	1.965	2.334	2.586	2.82	3.107	3.31
∞		1.282	1.645	1.96	2.326	2.576	2.807	3.09	3.291

The absolute value of the critical value depends on the confidence level (A), significance level (P), and the degree of freedom (DF), which is the sample size - 1 or $N - 1$.

What are the critical value(s) for when the significance level is 0.05 and the sample size approaches infinity? - **Answer**: 1.96 If the absolute value of our t-statistic is greater than our critical value, then we can reject the null hypothesis.

Finally, what is a **p-value**? - **Definition**: a **p-value** is the probability that a test statistic at least as significant as the one observed would be obtained assuming that the null hypothesis were true. The smaller the p-value, the stronger the evidence against the null hypothesis.

The **procedure of hypothesis testing** involves four steps: 1. Formulate the **null hypothesis** and the **alternative hypothesis**. 2. **Identify and compute a test statistic** that can be used to reject or fail to reject the null hypothesis. Check your assumptions! 3. **Compute the test statistic and p-value** 4. **Compare the p-value to an acceptable significance value, α and compare the test statistic to acceptable critical value(s)**. If p-value $\leq \alpha$ and the test-statistic $\geq$ |critical value|, then the observed effect is statistically significant, the null hypothesis is rejected, and the alternative hypothesis is valid.

2.1 Now, let's try it ourselves!

We will work with the diabetes dataset to learn how to apply the procedure of hypothesis testing to real world data. - This dataset contains 442 diabetes patients with data on age (AGE), sex (SEX), body mass index (BMI), mean arterial blood pressure (MAP), six blood serum measurements (TC, LDL, HDL, TCH, LTG, and GLU), and a quantitative measure of disease progression (Y).

```python
[1]: # Import pandas, so that we can import the diabetes dataset and work with the
     ↪data frame version of this data
     import pandas as pd
```

```python
[2]: # Set the path
     path = 'https://raw.githubusercontent.com/GWC-DCMB/curriculum-notebooks/master/'
     # This is where the file is located
     filename = path + 'SampleData/diabetes.csv'
```

```python
[3]: # Load the diabetes dataset into a DataFrame
     diabetes_df = pd.read_csv(filename)
     diabetes_df
```

```
[3]:      AGE  SEX   BMI     MAP   TC    LDL    HDL   TCH     LTG  GLU    Y
     0     59    2  32.1  101.00  157   93.2  38.0  4.00  4.8598   87  151
     1     48    1  21.6   87.00  183  103.2  70.0  3.00  3.8918   69   75
     2     72    2  30.5   93.00  156   93.6  41.0  4.00  4.6728   85  141
     3     24    1  25.3   84.00  198  131.4  40.0  5.00  4.8903   89  206
     4     50    1  23.0  101.00  192  125.4  52.0  4.00  4.2905   80  135
     ..   ...  ...   ...     ...  ...    ...   ...   ...     ...  ...  ...
     437   60    2  28.2  112.00  185  113.8  42.0  4.00  4.9836   93  178
     438   47    2  24.9   75.00  225  166.0  42.0  5.00  4.4427  102  104
     439   60    2  24.9   99.67  162  106.6  43.0  3.77  4.1271   95  132
```

```
440   36    1  30.0   95.00  201  125.2  42.0  4.79  5.1299   85  220
441   36    1  19.6   71.00  250  133.2  97.0  3.00  4.5951   92   57
```

```
[442 rows x 11 columns]
```

We are interested in understanding whether there are differences in diabetes progression by sex, i.e. is the disease progression different for males vs. females?

1. Formulate the null hypothesis and the alternative hypothesis. - **Null hypothesis**: There is NO difference in disease progression between male and female. - **Alternative hypothesis**: There is a difference in disease progression by sex.

```
[4]:  # Import numpy
      import numpy as np
```

```
[5]:  # Look at all unique values for sex
      np.unique(diabetes_df["SEX"])
```

```
[5]:  array([1, 2])
```

Males are indicated by "1" for the variable "SEX", while females are indicated by "2".

```
[6]:  # Define a vector of the disease progression for males and name it␣
      ↪progression_male
      diabetes_male = diabetes_df.query('SEX == 1')
      progression_male = diabetes_male['Y']

      # Define a vector of the disease progression for females and name it␣
      ↪progression_female
      diabetes_female = diabetes_df.query('SEX == 2')
      progression_female = diabetes_female['Y']
```

2. Identify and compute a test statistic that can be used to reject or fail to reject the null hypothesis. - As we are working with two independent samples, we will use the two-sample t-test and use the t-statistic.

3. Compute the test statistic and p-value.

```
[7]:  # Import stats methods to help calculate the t-statistic and p-value
      from scipy import stats
```

```
[8]:  # Run a Student's t-test with the method ttest_ind
      t_statistic, p_value = stats.ttest_ind(progression_male, progression_female)

      # Print out the test statistic and p-value
      print("t-statistic = " + str(t_statistic))
      print("p-value = " + str(p_value))
```

```
t-statistic = -0.9041147550244715
p-value = 0.3664292946519826
```

4. Compare the p-value to an acceptable significance value, α and compare the test statistic to acceptable critical value(s). If p-value $< \alpha$ and the test-statistic $\geq$ +critical value or test-statistic $\leq$ -critical value, that the observed effect is statistically significant, the null hypothesis is rejected, and the alternative hypothesis is valid.** - p-value $= 0.36 > 0.05$, so we fail to reject the null hypothesis. - t-statistic $= -0.90 > -1.96$, so this reaffirms that we fail to reject the null hypothesis. - Interpretation: There is no significant difference in diabetes progression between males and females.

Congratulations! You know how to conduct hypothesis testing with a Student's t-test!

Misconceptions about statistical significance: 1. A low p-values implies a large effect. - **Proper interpretation**: A low p-value indicates that the outcome would be highly unlikely if the null hypothesis were true. A lower p-value does not usually equate to a large effect. There are cases when a low p-value can occur with a small effect. 2. A non-significant outcome (AKA high p-value) means that the null hypothesis is probably true. - **Proper interpretation**: A non-significant outcome (AKA high p-value) means that the data do not conclusively demonstrate that the null hypothesis is false. This is why we should say, "When the p-value > 0.05, we fail to reject the null hypothesis." We should not say that we accept the null hypothesis when the p-value > 0.05.

Awesome work! You just learned about statistical significance! You learned:

- Important definitions such as **statistical significance, statistical significance level, null hypothesis, alternative hypothesis, test statistic, p-value, and critical values**.
- To **conduct hypothesis testing**.
- To **determine critical values** to compare with your own test statistic in order to decide whether a variable has an effect on the outcome of interest.
- To implement a Student's t-test.